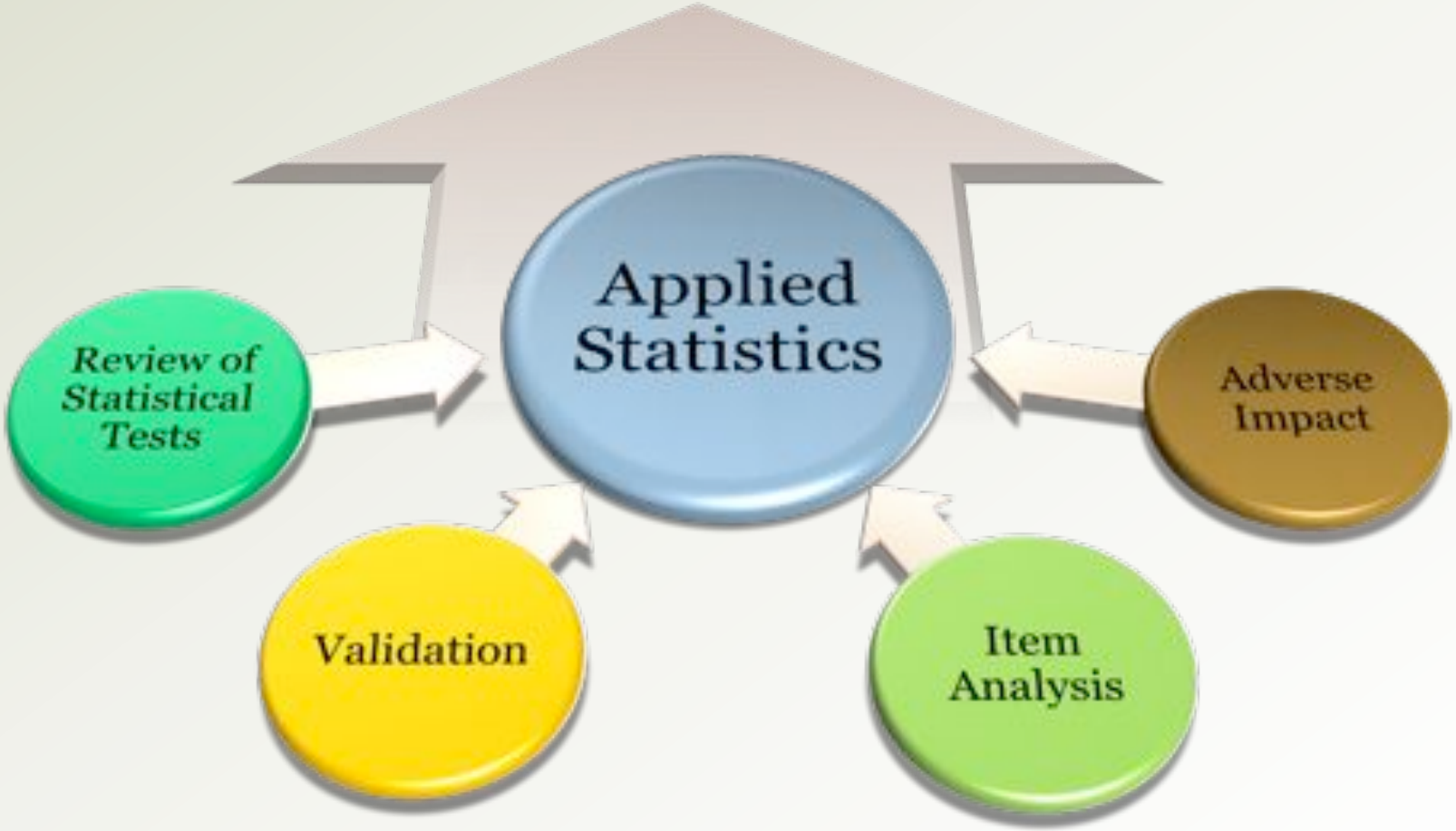
An Introduction to Applied Statistics For HR Professionals

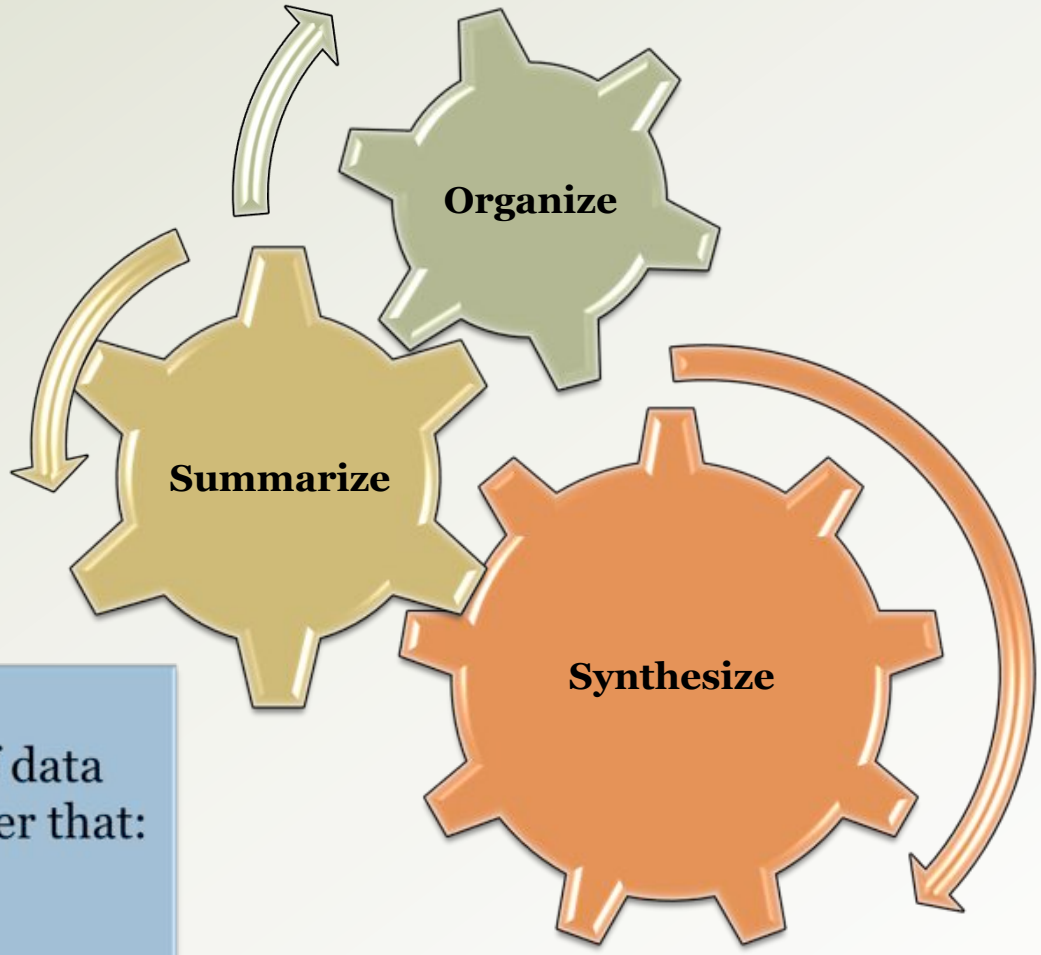
Jim Higgins, Ed.D.
www.BCGinstitute.org

Overview of Presentation

Don't Just Survive. Thrive!



What Are Descriptive Statistics



The Goal:
To take large amounts of data and present it in a manner that:

- Is easy to understand
- Tells "the story"
- Communicates the KEY point

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Which Is Easier To Understand?

Gender	Age	Test Score
Male	35	78
Male	28	73
Female	41	82
Male	40	77
Female	33	85
Female	38	77
Female	30	80
Male	21	69
Male	29	74
Male	27	79
Male	34	70
Female	31	78
Male	36	72
Male	33	77

- Sample Size = 14
- Gender
  - Male = 9 (64.3%)
  - Female = 5 (35.7%)
- Average Age = 33
  - Males = 31.4
  - Females = 34.6
- Average Test Score = 76.5
  - Males = 74.3
  - Females = 80.4

Data vs. Information!

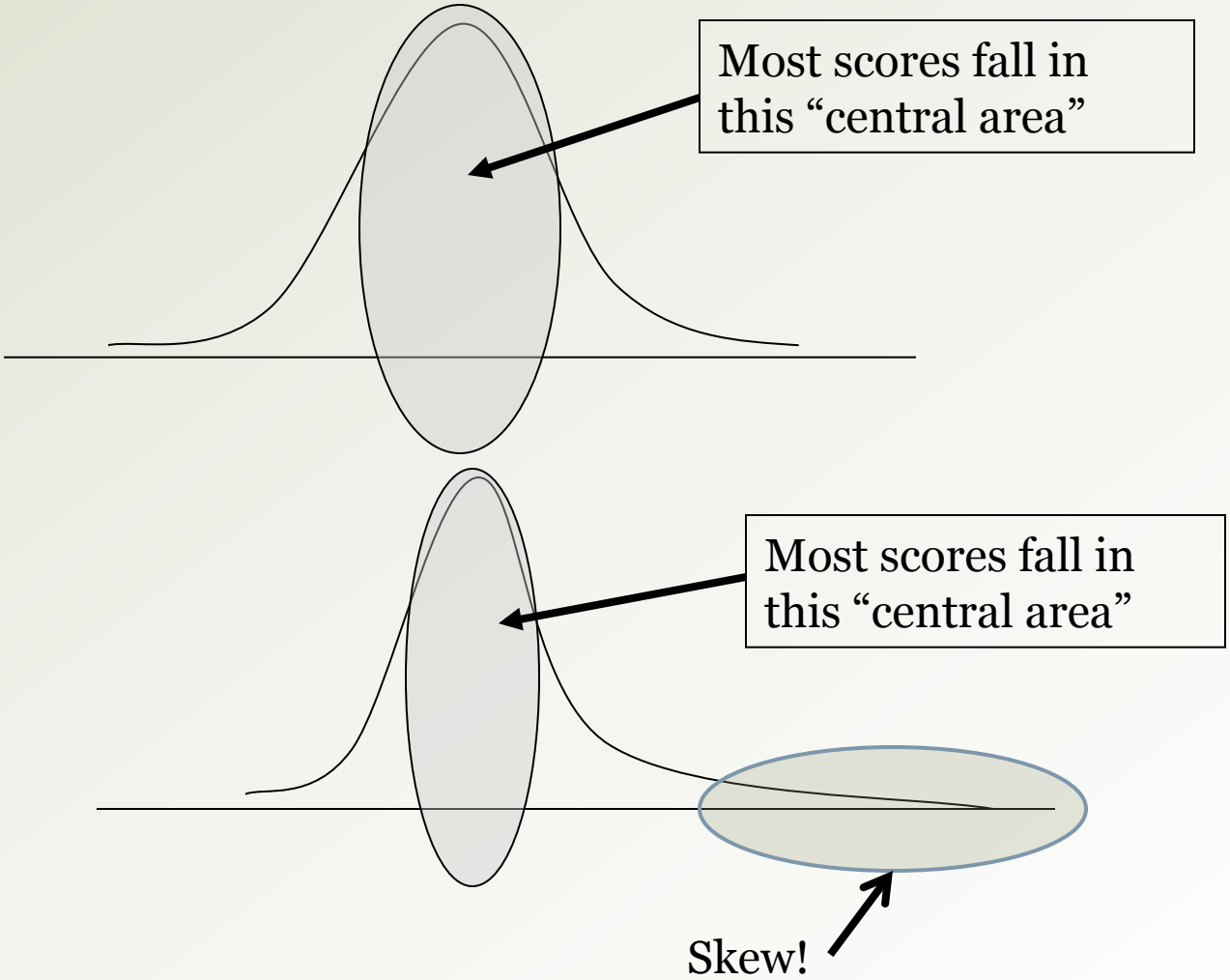
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

What is a Measure of Central Tendency?

- A single summary number that tells you about what is **most common**.
- Gives you an idea about what is the most common number in a distribution of scores.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Examples of Central Tendency



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Summary: Measures of Central Tendency

There are 3 measures of central tendency:

- Mode: Most Frequently occurring number
- Median: Number where 50% are higher and 50% are lower
- Mean: The arithmetic "Average"

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Important!

Each measure of central tendency is looking at what is “most common” about a distribution but…

Each takes a different perspective!

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Check This Out!

- If we have five people aged 12, 12, 15, 17, 45
  - What is the Mode?
  - What is the Median?
  - What is the Mean?

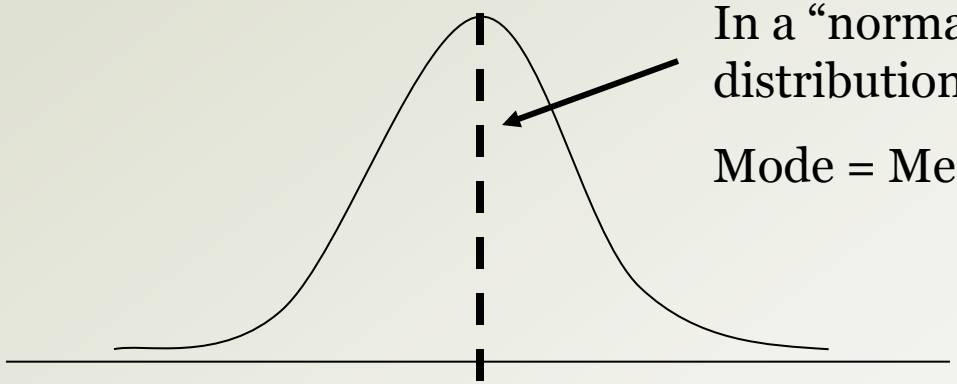
The Mode is…12

The Median is…15

The Mean is…20.2

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

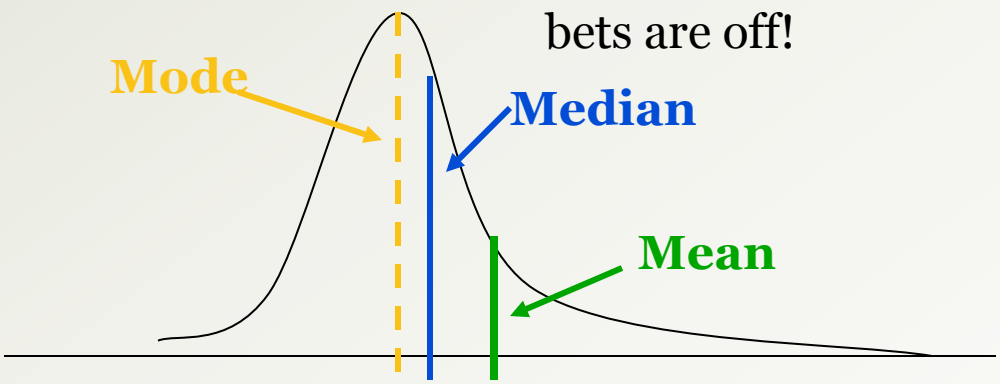
A Picture is Worth…



In a "normal" distribution:

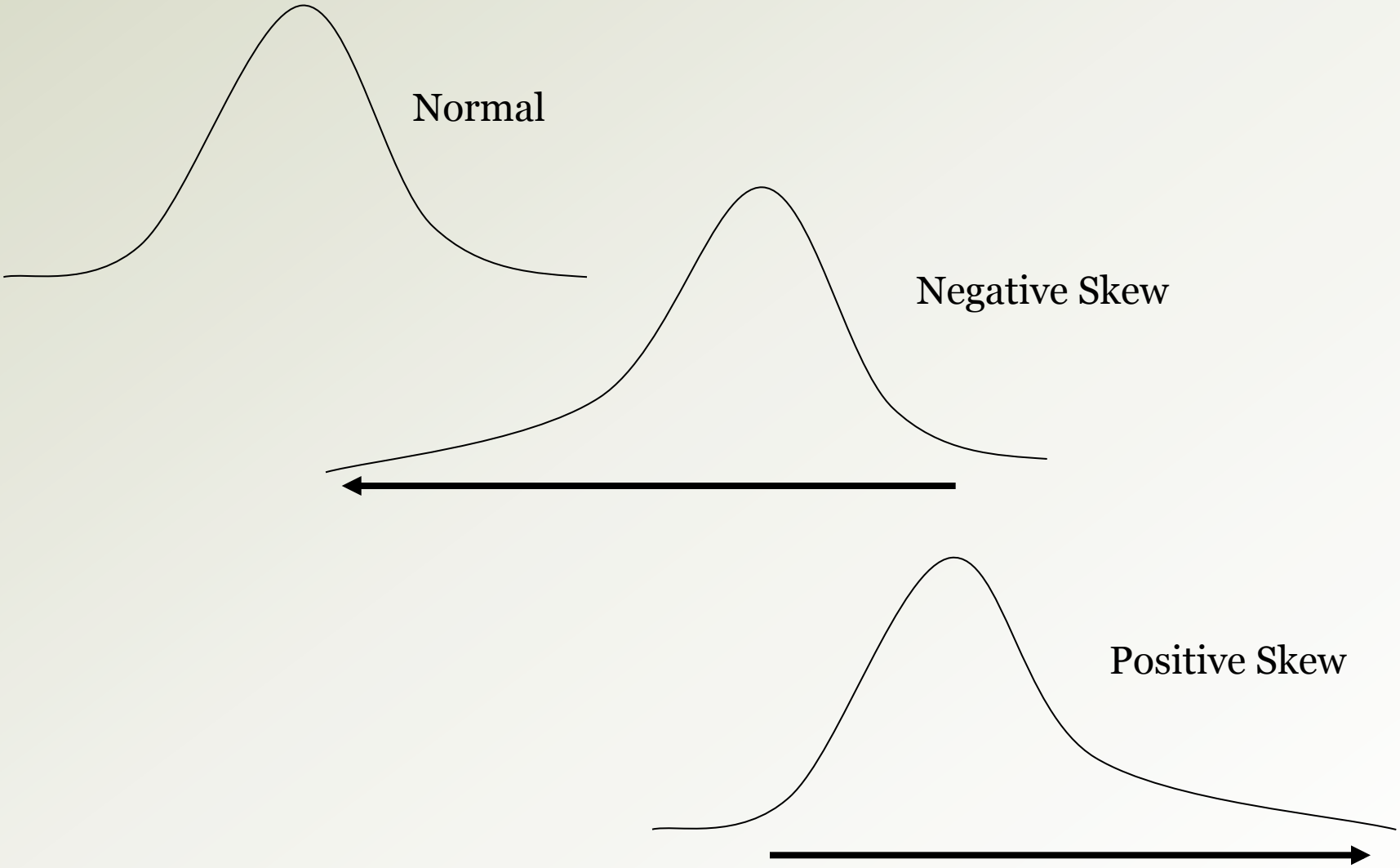
Mode = Median = Mean

In a "Skewed" distribution, all bets are off!



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Skewness



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

So What?

- The Mode ALWAYS means the same thing… the most common score.
- The Median ALWAYS means the same thing… 50% higher and 50% lower.
- The Mean ONLY MAKES SENSE WHEN YOU HAVE A NORMAL DISTRIBUTION. This has significant implications to interpreting test and other statistics!

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

What is Variation?

- Refers to how “spread out” scores are around the mean
- Tells you whether most scores are close to the mean or if they are really a lot different
- Small variation means most scores are close to the mean
- Large variation means many of the scores are much higher or much lower than the mean

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Four Important Measures of Variation

Least Important

- Range
- Sum of Squares
- Variance
- **Standard Deviation**

Most Important

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Sum Of Squares

$$\sum \left(X - \overline{X}\right)^2$$

In English:

The sum of the squared deviation scores

I Still Don't Get It:

The sum of the squared differences of each score from the mean

O.K., I may be dense…but I still don't get it!

Subtract the mean from each raw score and square the result. Once you have done it for each score, add them all up.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Breakin' Down The Sum Of Squares

$$X - \overline{X}$$

Subtracting the mean from a raw score results in a "deviation score". This tells you how much that score "deviates" from the mean.

$$\left(X - \overline{X}\right)^2$$

When you square a deviation score, you get…"a squared deviation score"!

$$\sum\left(X - \overline{X}\right)^2$$

When you add up the squared deviation scores, you get…"The sum of the squared deviation scores". The **Sum of Squares**!

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Calculating the Sum of Squares

Salary	Deviation Scores	Squared Deviation Scores
$10	-6.8	46.24
$12	-4.8	23.04
$13	-3.8	14.44
$24	7.2	51.84
$25	8.2	67.24
$\bar{X}$ = **16.8**	**Σ = 0**	**Σ = 202.8**

This is the "Sum of Squares" or, the sum of the squared deviations of each score from the mean.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Problems With The Sum of Squares

- The sum of squares suffers from the same problems as the range.
  - Hard to interpret in a meaningful way
  - Difficult to make comparisons between the variability of different distributions based on the sum of squares.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Variance

$$sd^2 = \frac{\sum (X - \overline{X})^2}{n - 1}$$

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Variance (Continued)

Sum of Squares

$$sd^2 = \frac{\sum (X - \overline{X})^2}{n - 1}$$

The number of people in your set of data "minus 1"

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Calculating the Variance

Salary	Deviation Scores	Squared Deviation Scores
$10	-6.8	46.24
$12	-4.8	23.04
$13	-3.8	14.44
$24	7.2	51.84
$25	8.2	67.24
$\bar{X}$ = **16.8**	**Σ = 0**	**Σ = 202.8**

Sum of Squares

Variance = 202.8 / 4

202.8 / 4 = 50.7

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

What is the variance?

- Point 1 – The sum of squares is the sum of the squared deviations of each score from the mean.
- Point 2 – The variance is the sum of squares divided by "n"
- Therefore, the variance is the average of the squared deviations of each score from the mean.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Problems with the Variance

- Generally, the same problems as the Range and the Sum of Squares
  - Hard to interpret
  - Hard to make comparisons
- Another important problems is that it is reported in "squared" units *(because you had to square deviation scores to get it)*

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Standard Deviation

The square root of the variance

$$sd = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

When you take the square root of the variance, you change the variance from being squared deviation scores back into raw scores.

The standard deviation tells you, on average, how much scores deviate from the mean.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Standard Deviation

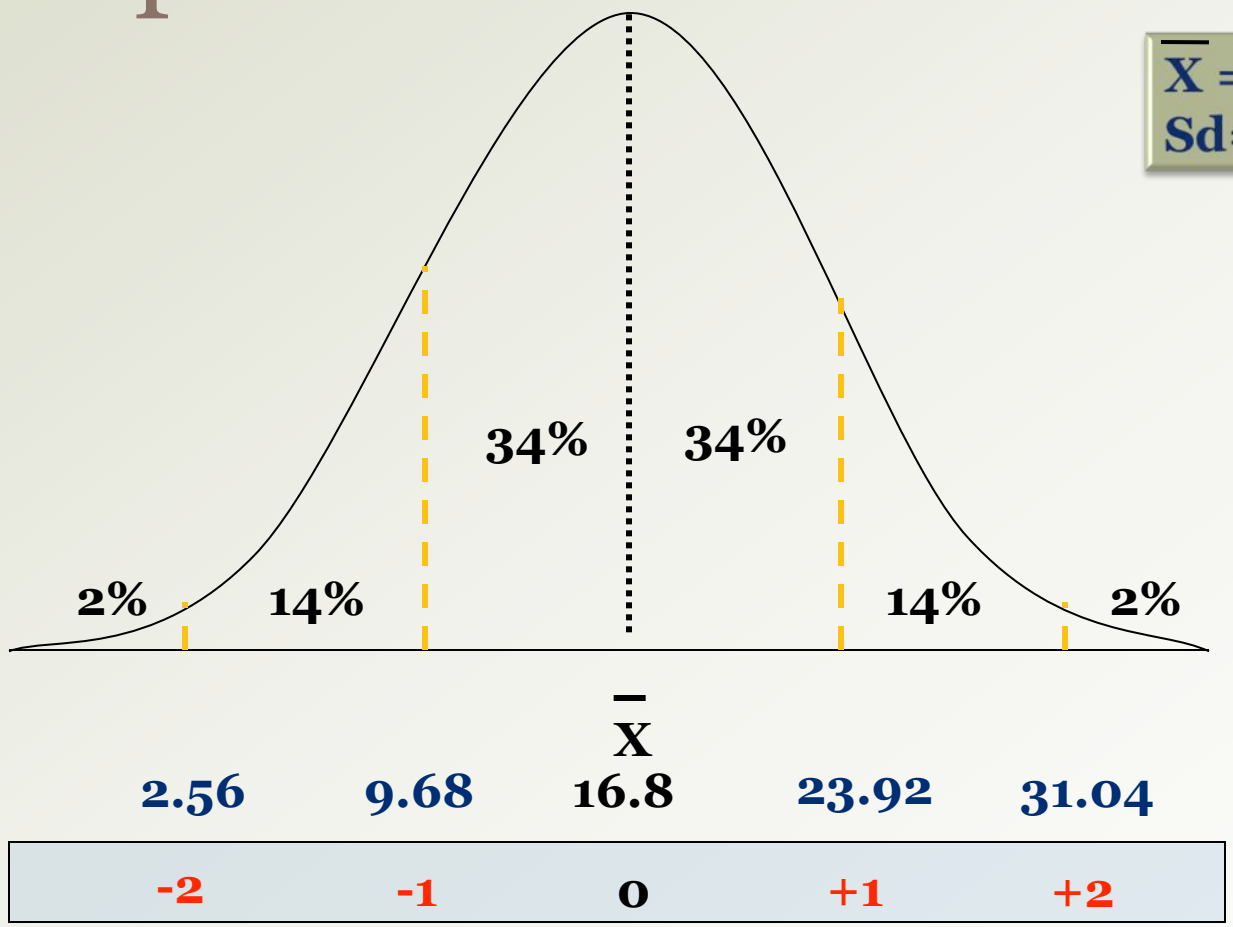
The square root of the variance

$$sd = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}}$$

The standard deviation tells you, on average, how much scores deviate from the mean.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Example



The Z-Distribution

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Summary

- We talked about 4 measures of variation
  - Range
  - Sum of Squares
  - Variance
  - Standard Deviation
- The most important and useful of these is the Standard Deviation

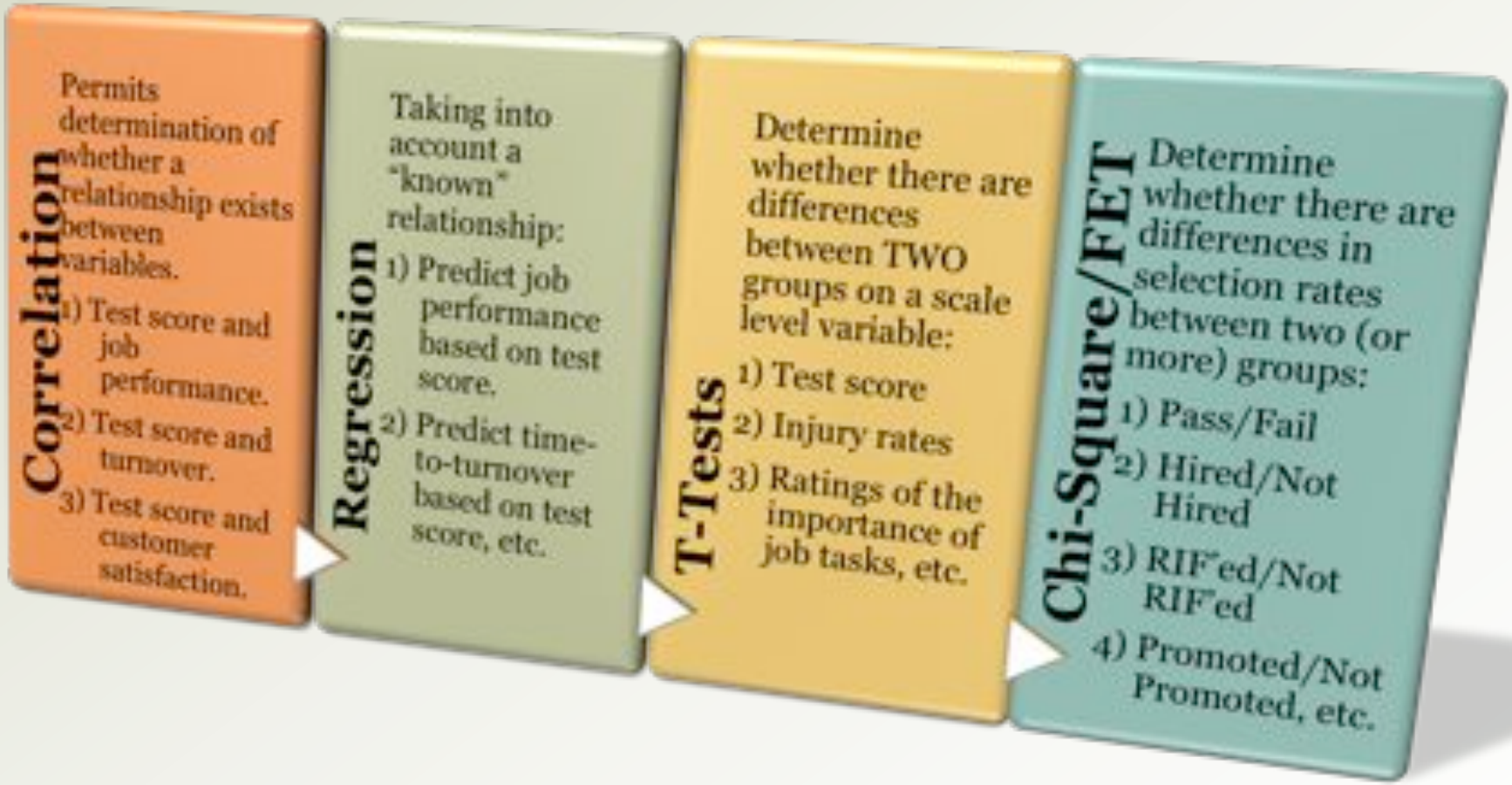
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Consider the Implications!

- The mean is sensitive to "skewness." It is pulled in the direction of the skew.
- In a skewed distribution, the mean may not even effectively measure the distribution's central tendency (i.e., tell you what is most common).
- Since the mean is used to calculate most other important statistics, if the mean is not accurate, then all statistical results may be open to question.
- The importance of understanding the mean and standard deviation cannot be overstated!

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

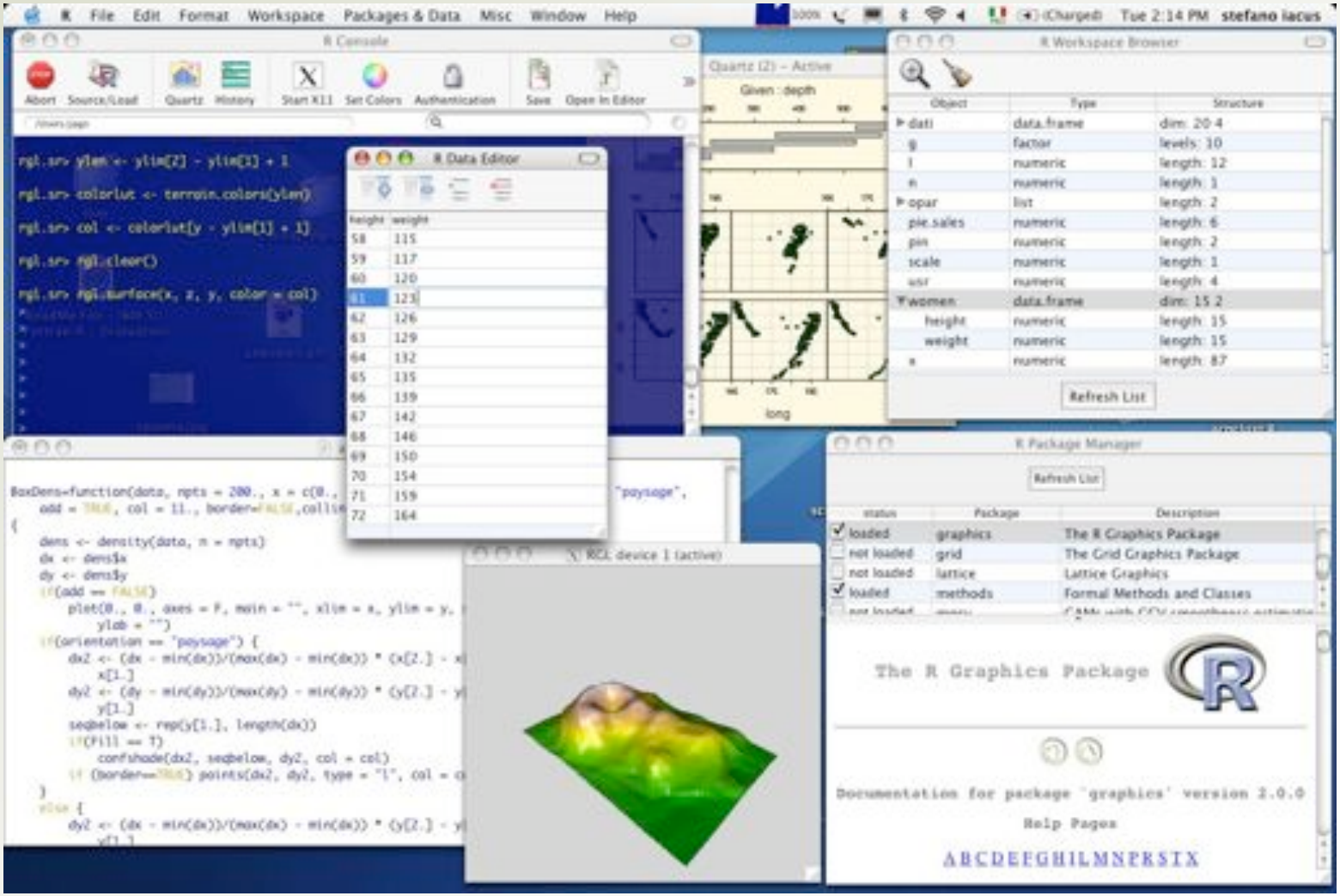
Important Statistical Tests



Each Test Has Its Role

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Validation


File Edit Format Workspace Packages & Data Misc Window Help
Tue 2:14 PM stefano iacus
R Console
R Data Editor
R Workspace Browser
R Package Manager
Refresh List
The R Graphics Package
Documentation for package 'graphics' version 2.0.0
Help Pages

What Needs To Be Validated?

Any Practice, Procedure or Test that has adverse impact and which is used in a manner that screens people into pass/fail, promote/not promote, select/ not select, etc. groups.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Test Validation Is All About Showing That Your Test Is…

Uses correlation.

Reliable

As unbiased as possible

T-test, Chi, FET.

Job Related

Content validation relies on getting **Average** Importance, Frequency, When Required ratings from SMEs.
Criterion validation relies on demonstrating statistical relationship between test and criterion using a **Correlation Coefficient**.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

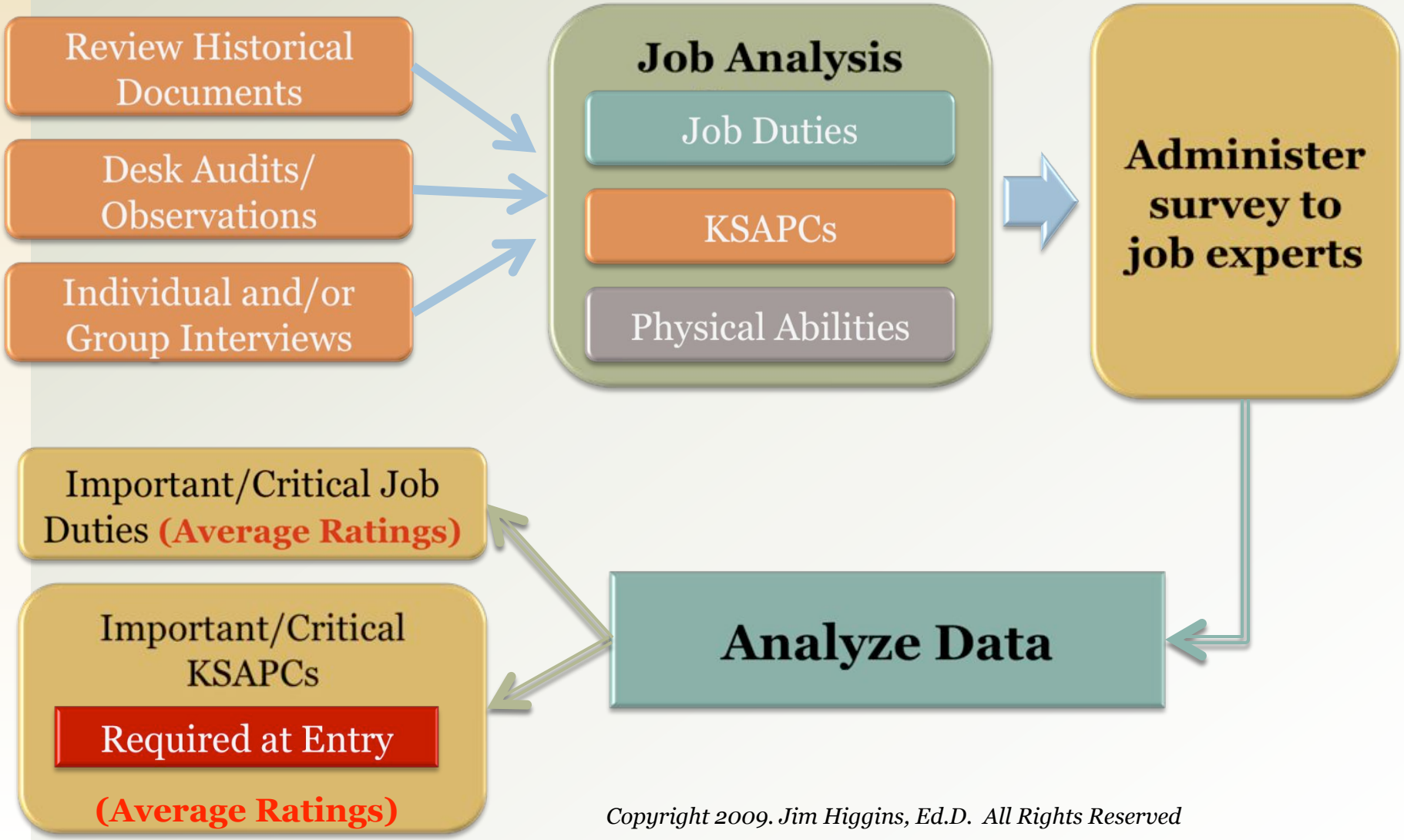
What Types of Validation are REALLY USED in Practice?

- **Content** (80%)
- **Criterion** (15%)
- **Construct** (5%)

  - **Why?**
    - *Simplicity*
    - *Defensibility*

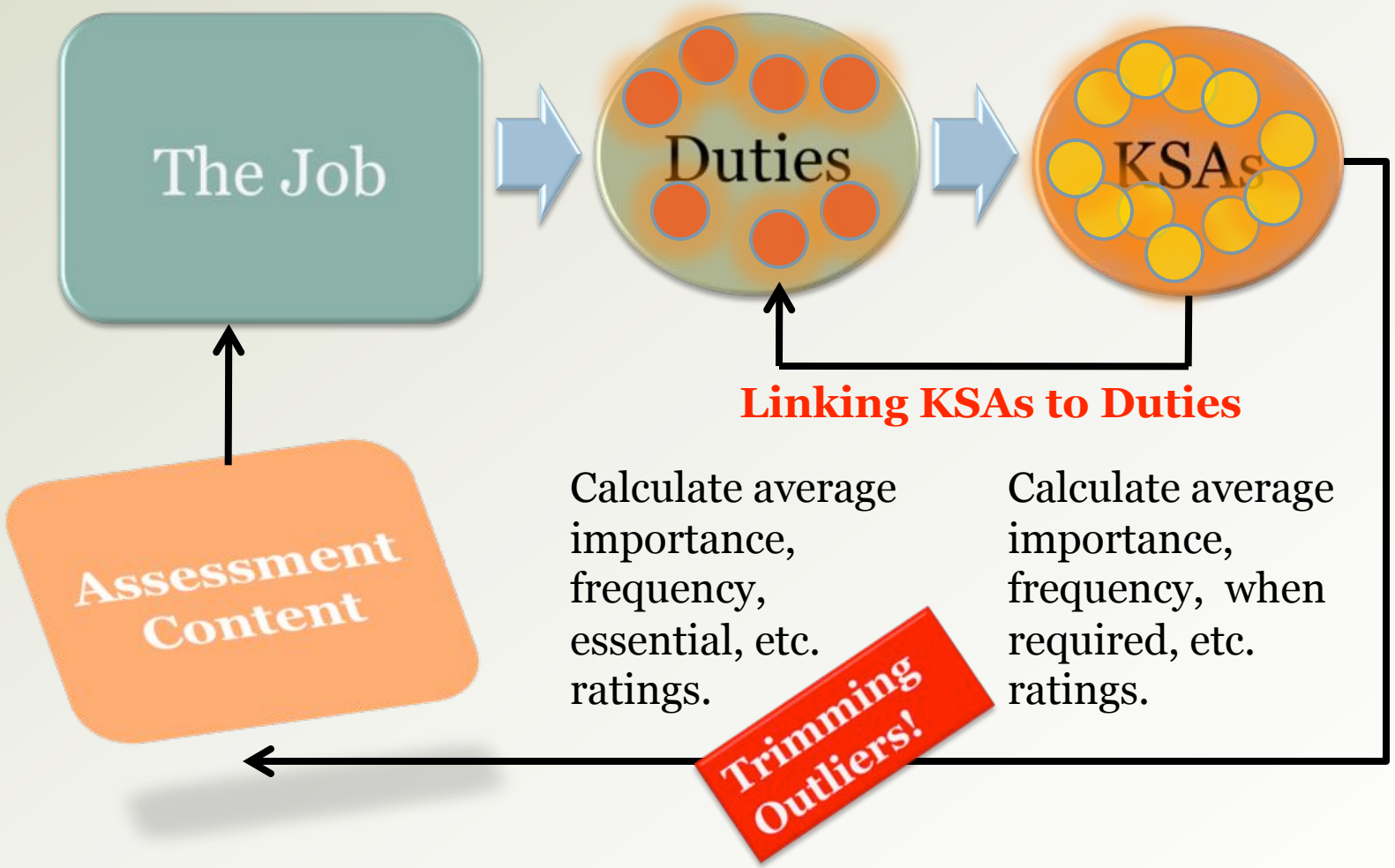
Copyright 2007 Biddle Consulting Group
Used by Permission

Job Analysis: The Foundation of Content Validation



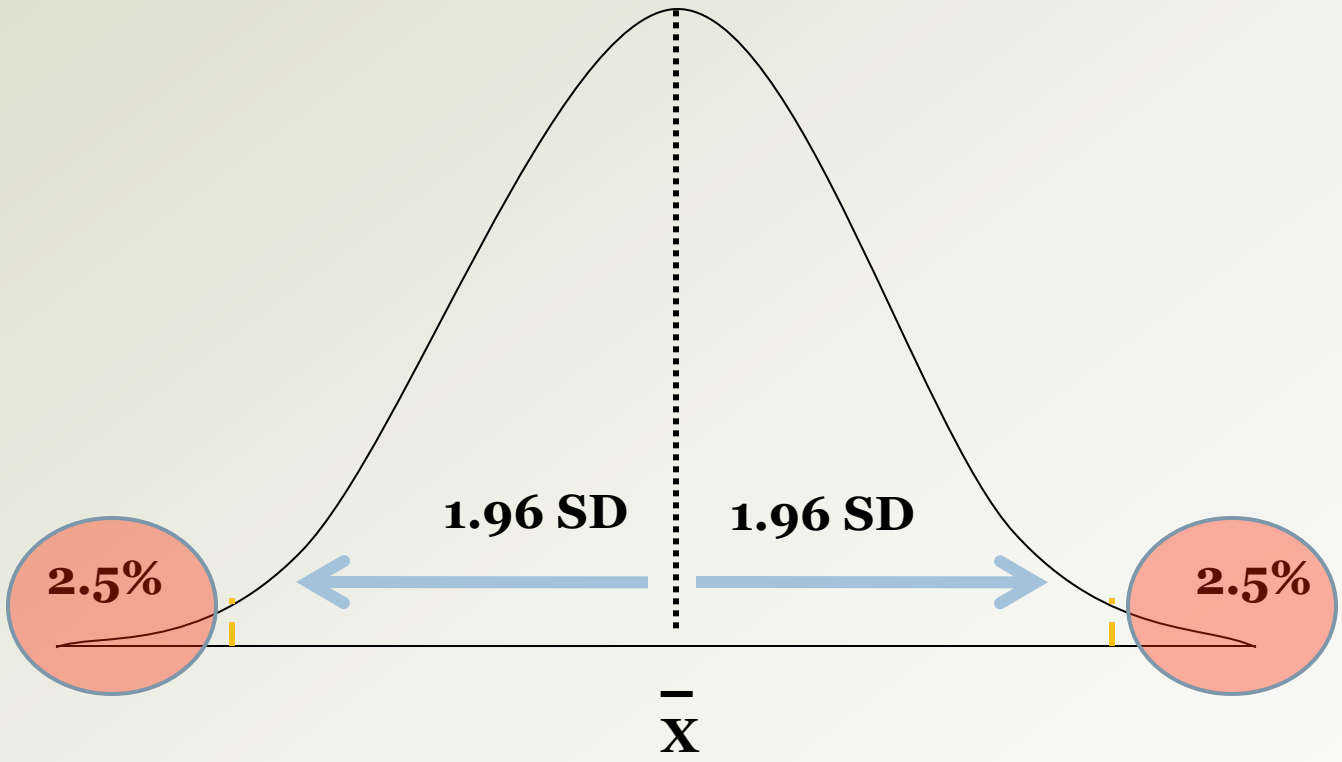
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Test Development Process



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

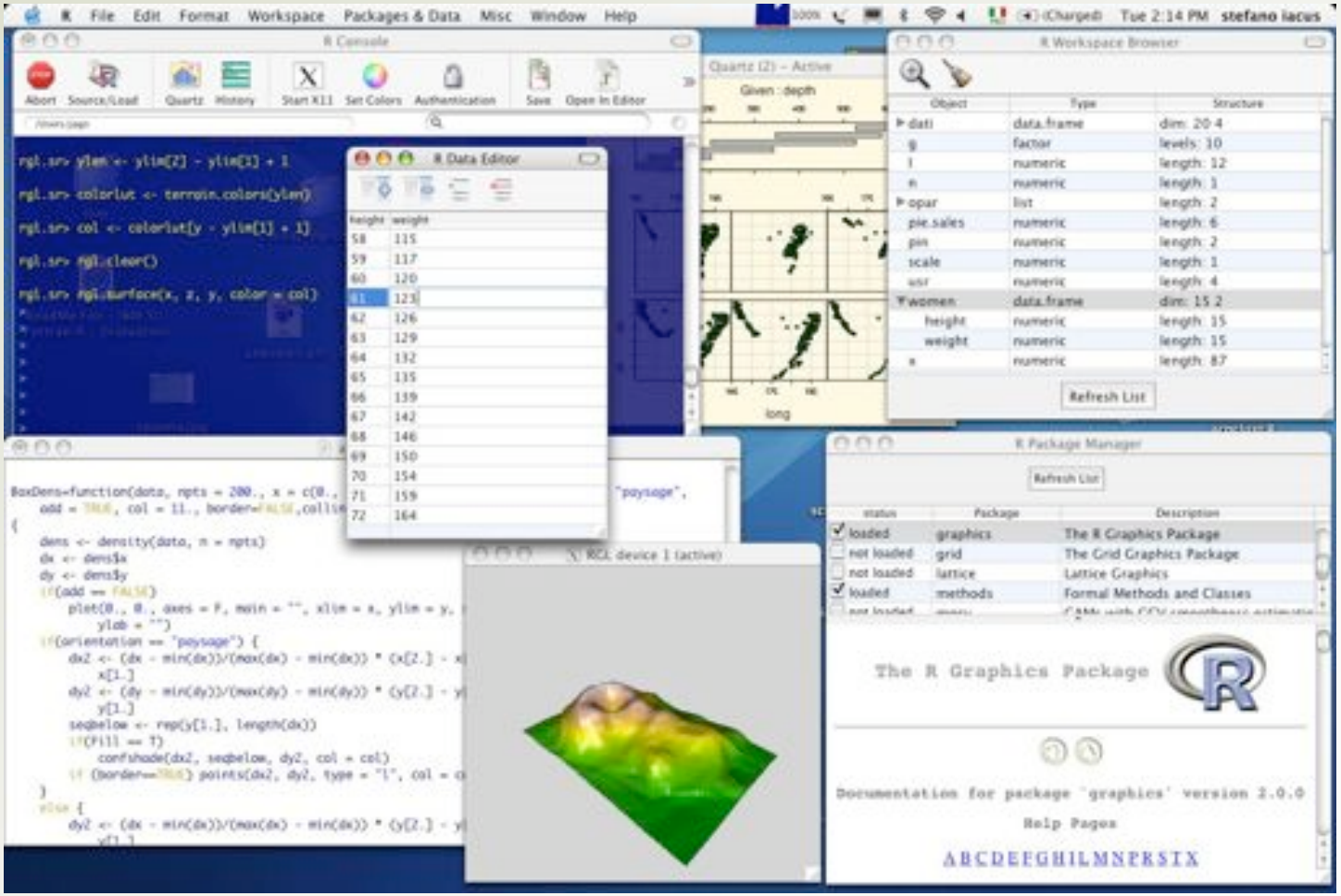
95% Trimmed Mean



Step 1: Calculate the mean
Step 2: Calculate the SD
Step 3: Eliminate cases whose rating is greater than 1.96 SD above/below the mean.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Item Analysis



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Optimizing A Test: Item Analysis

Effective Items

Distractor Analysis

Do high scorers choose the "Key" while low scorers chose distractors?

Are there disctractors that nobody is choosing?

Item-Total Correlation

Negative: High scorers perform poorly

None: Item is useless

Positive (>.20): High scorers tend to answer correctly

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Example Item Analysis Statistics: Distractor Analysis

Item 1 Grouping	A	B	C	D*
Highest 1/3	10%	20%	15%	55%
Middle 1/3	15%	20%	40%	25%
Lowest 1/3	25%	20%	35%	20%
Total	10%	15%	25%	45%

Item 2 Grouping	A	B*	C	D
Highest 1/3	35%	34%	10%	21%
Middle 1/3	10%	50%	34%	16%
Lowest 1/3	0%	45%	20%	35%
Total	25%	20%	30%	25%

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Example Item Analysis Statistics: Item-Total Correlation

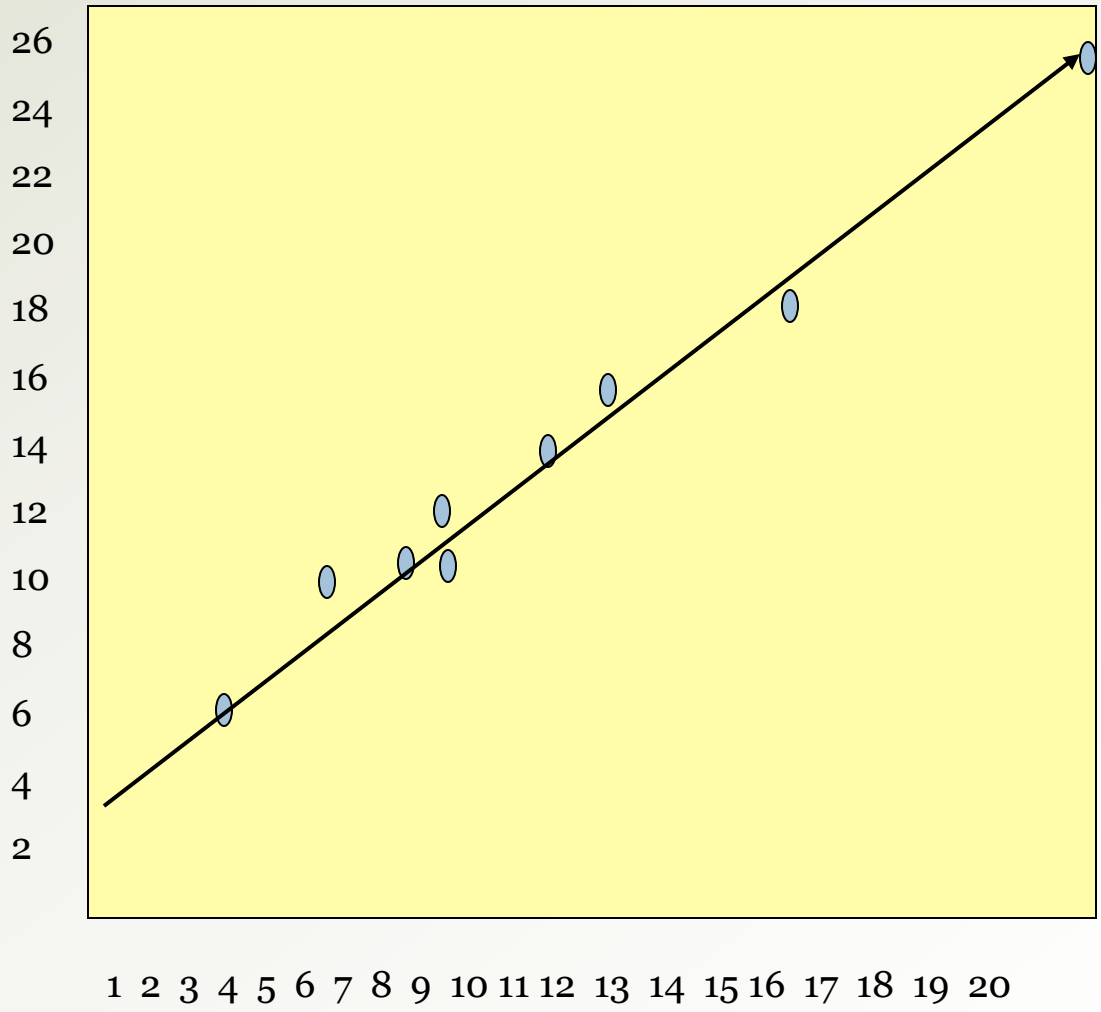
Item 1 Grouping	A	B	C	D*
Highest 1/3	0.00	0.12	-0.15	0.35
Middle 1/3	0.03	-0.24	0.05	0.24
Lowest 1/3	-0.04	-0.31	0.10	0.26
Total	-0.12	0.09	0.11	0.28

Item 2 Grouping	A	B*	C	D
Highest 1/3	0.41	-0.21	0.00	0.05
Middle 1/3	***	0.10	0.03	0.15
Lowest 1/3	0.20	-0.05	-0.14	-0.10
Total	0.32	-0.18	-0.12	0.01

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

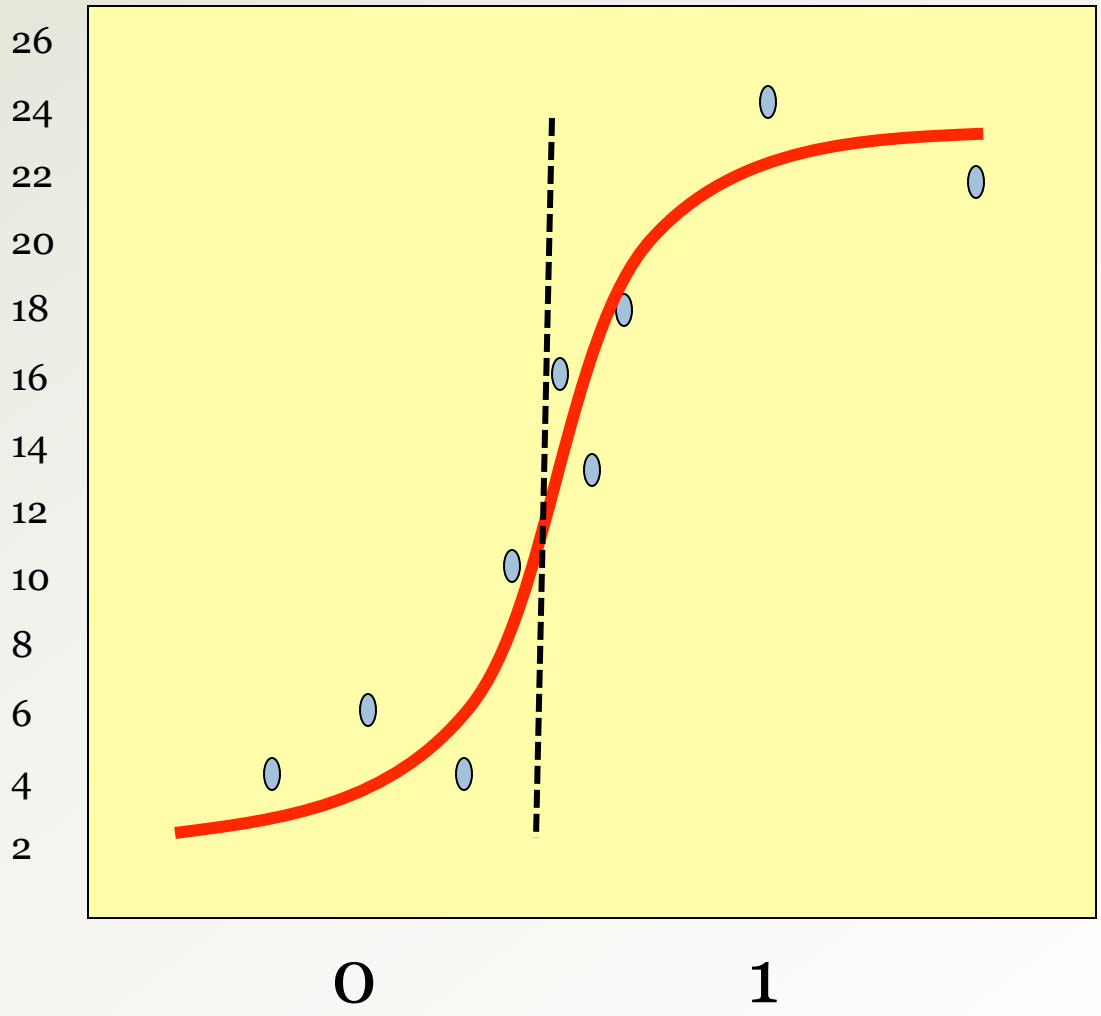
Test-Criterion Correlations

Test Score	Productivity
4	6
6	10
9	12
11	14
9	10
12	16
8	10
15	18
20	25



Item-Total Correlations

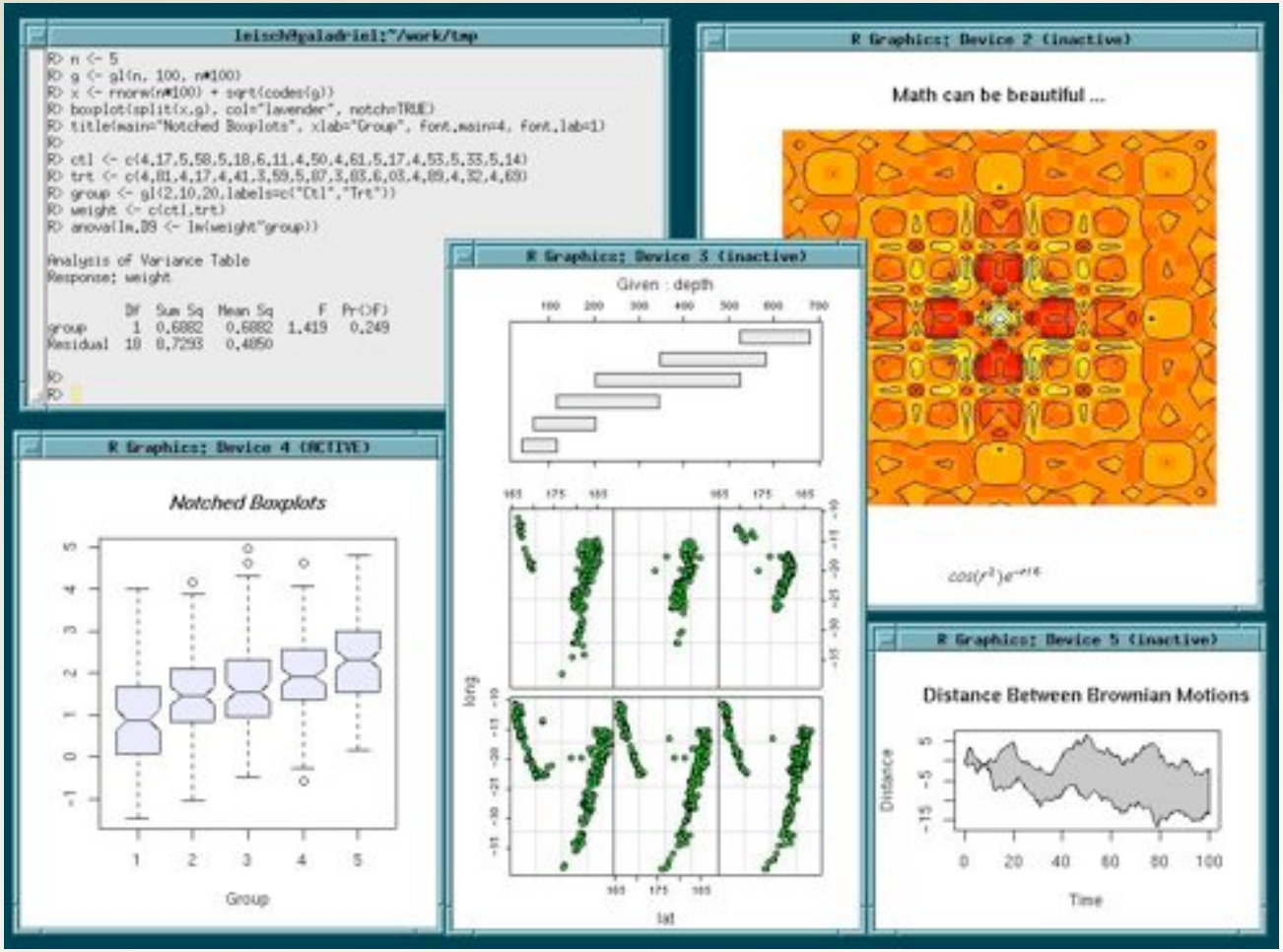
Item Score	Test Score
0	6
0	10
0	12
0	14
1	10
1	16
1	10
1	18
1	25



What Item-Total Correlations Tell You About Test Items

- Positive item-total correlations tell you that people who answer the item correctly tend to get higher test scores than those who don't. This would be a GOOD THING!
- Negative item-total correlations tell you that people who answer the item incorrectly tend to get lower scores than those who don't. This is "generally" a BAD THING.
- A "zero" (r=0.00) item-total correlation means that the test item does not provide any additional information about how well a person does on the test. This "generally" is a BAD THING.
- The bigger the "positive" item-total correlation you have, the better the item—as long as the distractors are working correctly.
- Generally you want to see item-total correlations of at least r=0.20.

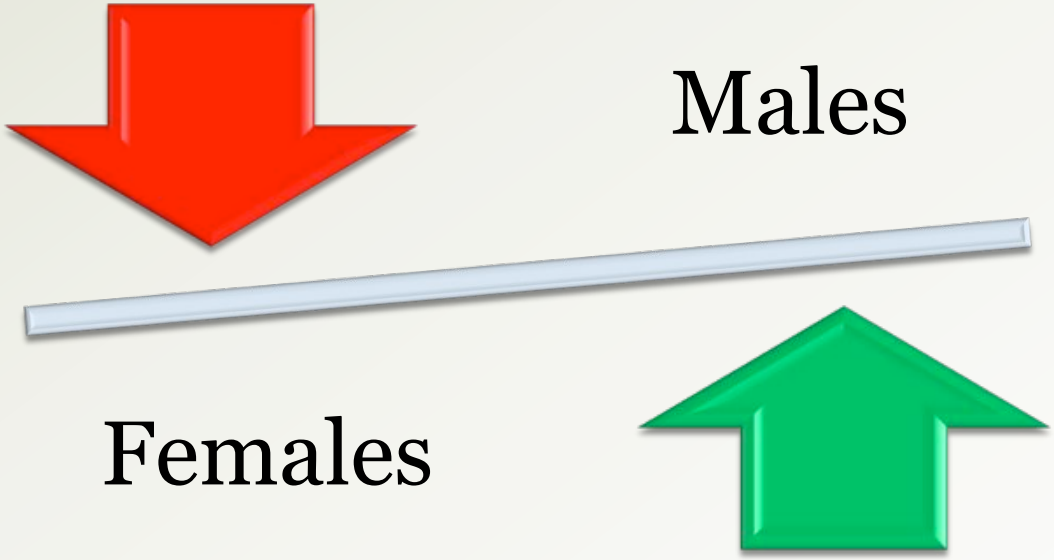
Adverse Impact



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Selection Rate Comparison

Evaluates whether or not a practice, procedure or test is disproportionately impacting a sex or racial group. If the individual practice, procedure or test cannot be separated and analyzed on its own, the entire selection process may be analyzed to evaluate adverse impact.

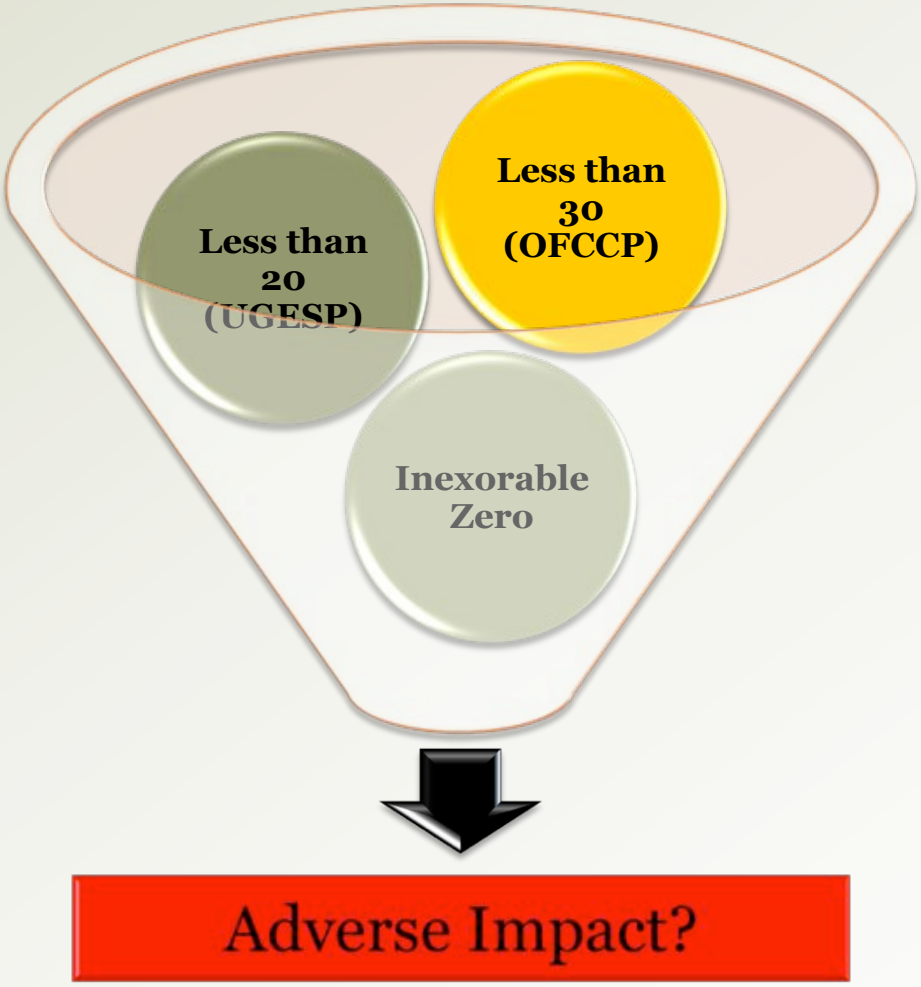


Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

A Note About Sample Size

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

When is a Sample Size Considered Too Small for Adverse Impact Analysis



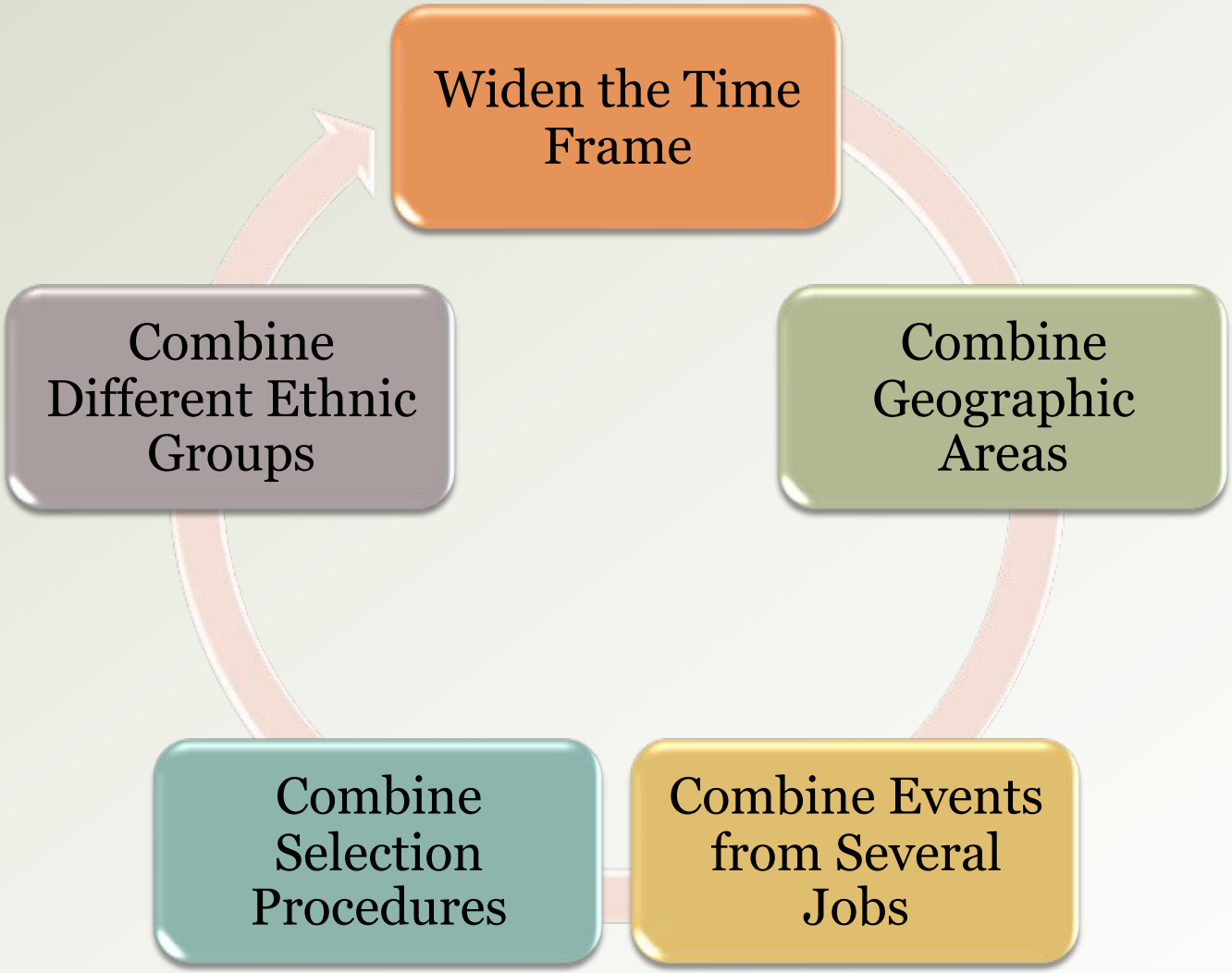
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Courts Are Not Unanimous On Issues Related To Sample Size

- Supreme Court warns us to be wary of using simplistic percentage comparisons based on small sample sizes.
- However, to appear legitimate, you may be forced to consider ways to increase your sample size to permit a statistical analysis
  - Increasing sample size is more "plaintiff friendly"
  - Avoiding artificially inflating sample size is more "employer friendly"

Copyright 2007 Biddle Consulting Group
Used by Permission

Ways to Increase Sample Size



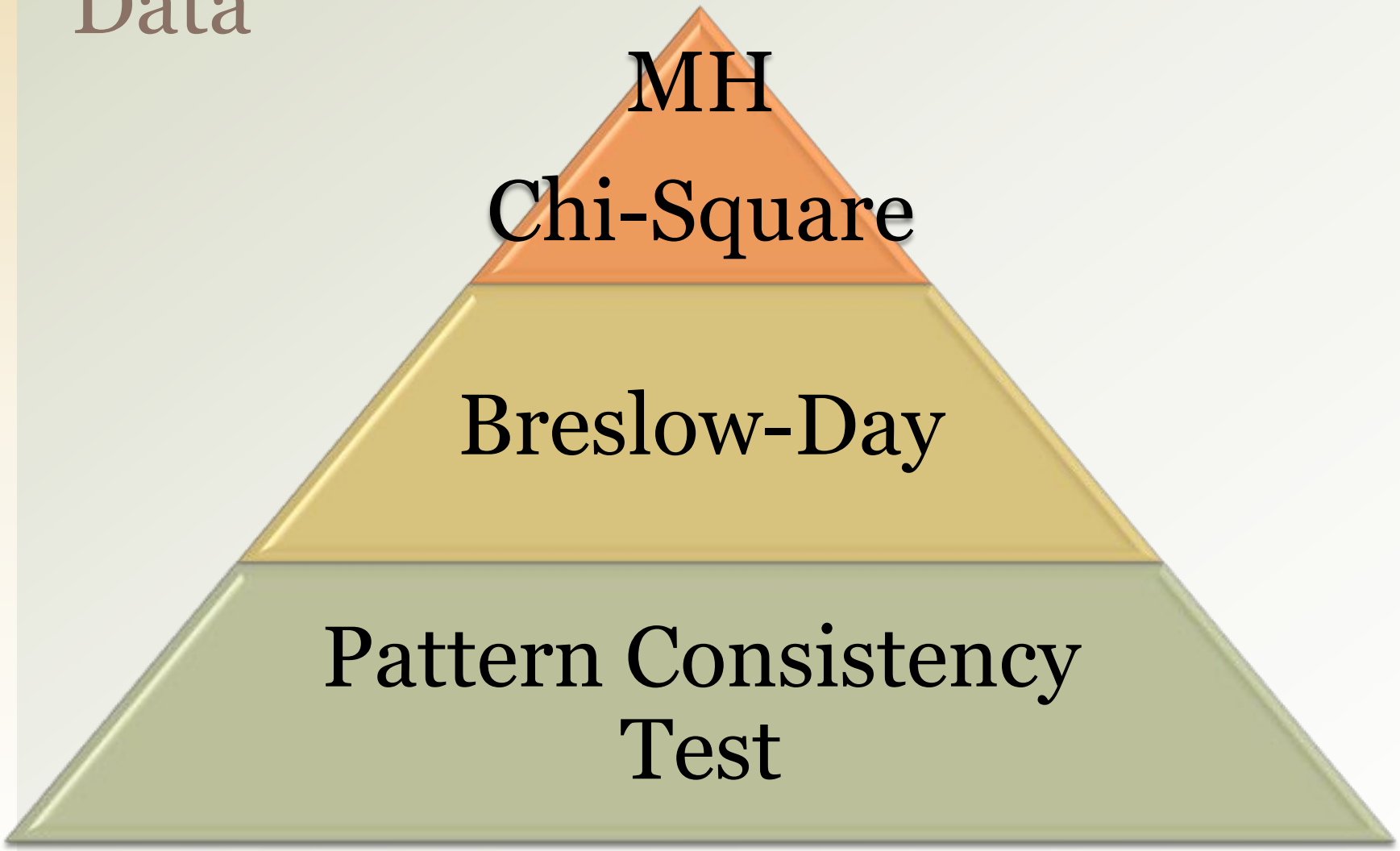
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Simpson's Paradox

Year	Group	# Applicants	# Selected	Selection Rate
2005	Men	400	200	50%
	Women	100	50	50%
2006	Men	100	20	20%
	Women	100	20	20%
2005 + 2006	Men	500	220	44%
	Women	200	70	35%

Copyright 2007 Biddle Consulting Group
Used by Permission

When It Is Necessary To Combine Data



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Comparing Group Performance

- A t-test is a statistic that is designed to compare the scores of two groups (e.g., the test score received by males and females.)
- Passing rates (counts of the number of people passing a test or procedure) are calculated using a Chi-Square or FET.

How t-Tests Work (In a nutshell)

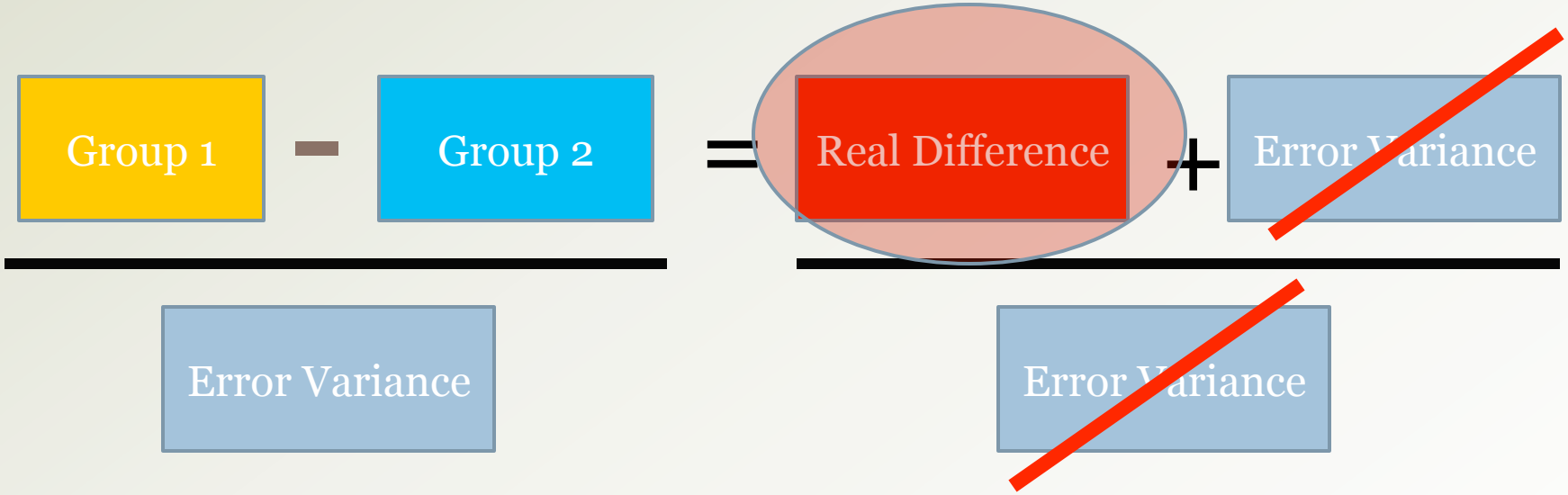
Group 1 − Group 2 = Difference

The difference, however, is composed of BOTH real differences AND differences that are simply due to random chance.

Difference = Real Differences + Error

How t-Tests Work (In a nutshell)

What is needed is a way to "filter out" the error part of the difference so that you have a measure of "real" differences that exist between groups.



The Problem With The t-Test Approach

- Too simplistic
- Fails to take into account that real differences may exist for legitimate reasons
- Results in findings of pay disparities when none exist

On The Other Hand

- Very robust statistical test
- Highly sensitive
- Works with small samples

The Problem With Chi Square Tests and the FET

- Chi Square works great on large sample sizes but can be biased with small sample sizes.
- All cells in a Chi Square need to have at least 1 person. If any cell has and expected frequency of 0 or if 20% or more of the cells have expected frequencies of less than 5 people, a special correction is applied called the "Yates Correction for continuity." Software such as SPSS calculates this for you.
- The FET is very powerful with small samples. It calculates the "exact probability" of seeing a given event. However, because the data used in such an analysis are not continuous, the FET can be overly conservative and fail to find statistical significance when there really is a difference in selection rates between groups.
- BCG recommends using the FET with a correction applied called the "Mid P Correction" (See www.BCGinstitute.org for more information or email the presenter at Jhiggins@biddle.com)

An Advanced Technique: Binary Logistic Regression

- Just because there is adverse impact, it does not mean that anything illegal is going on.
- If the selection procedure is valid but there are real differences in subgroup qualifications, it may be perfectly acceptable.
- Binary Logistic Regression is an advanced statistical procedure that may be used to control for legitimate factors that legitimately affect hiring, promotional, and termination procedures to determine if a persons gender or race play a role (i.e., discrimination)
- Can be useful for step analyses.

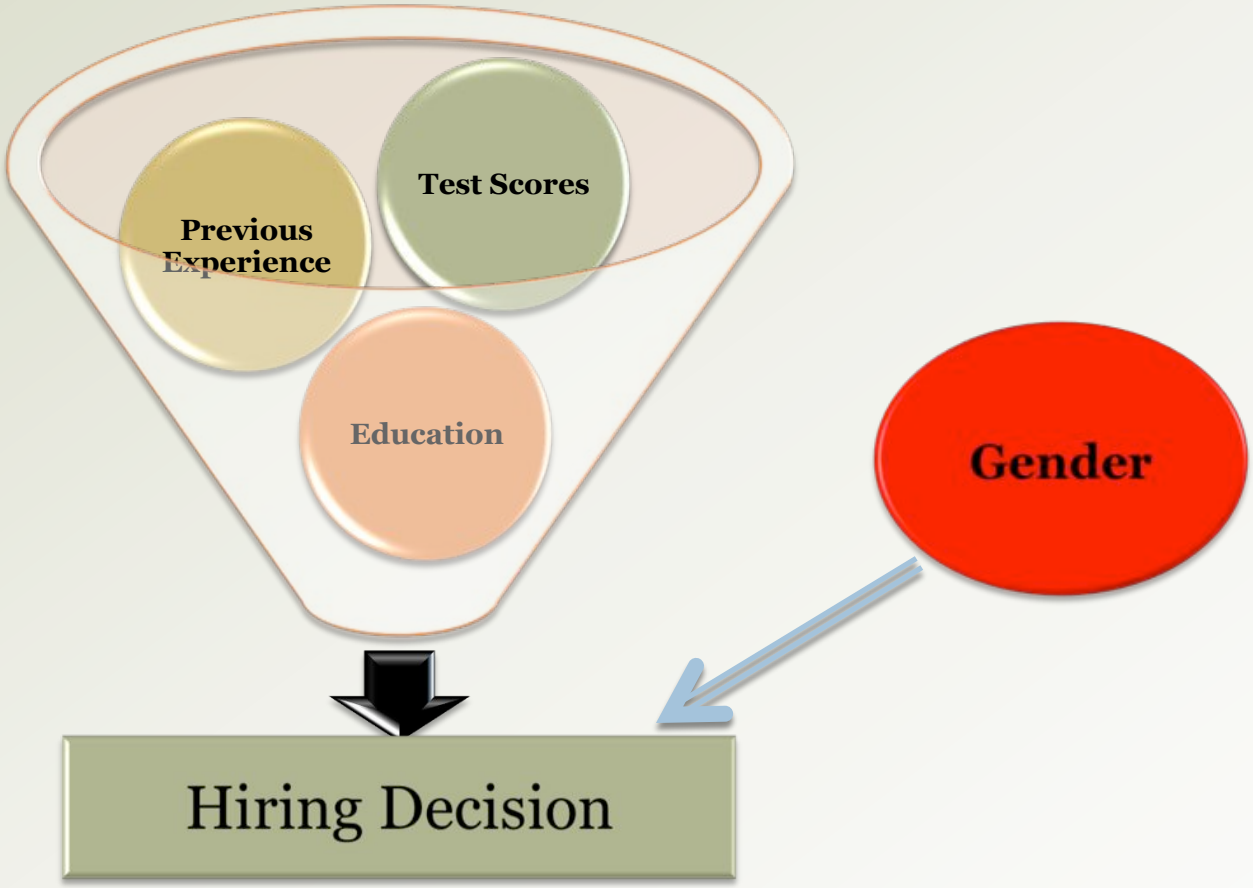
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

A Closer Look At Binary Logistic Regression

- Designed to "Model" or "Predict" dichotomous outcomes
  - Hired vs. Not Hired
  - Promoted vs. Not Promoted
  - Terminated vs. Not Terminated
  - Turned Over vs. Not Turned Over
  - Customer Purchase vs. Customer Not Purchased
- This VERY powerful tool can help you understand the underlying causes of such outcomes, how likely they are to occur, and whether one variable—after controlling for all others specified in the model—has an impact on the outcome.

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

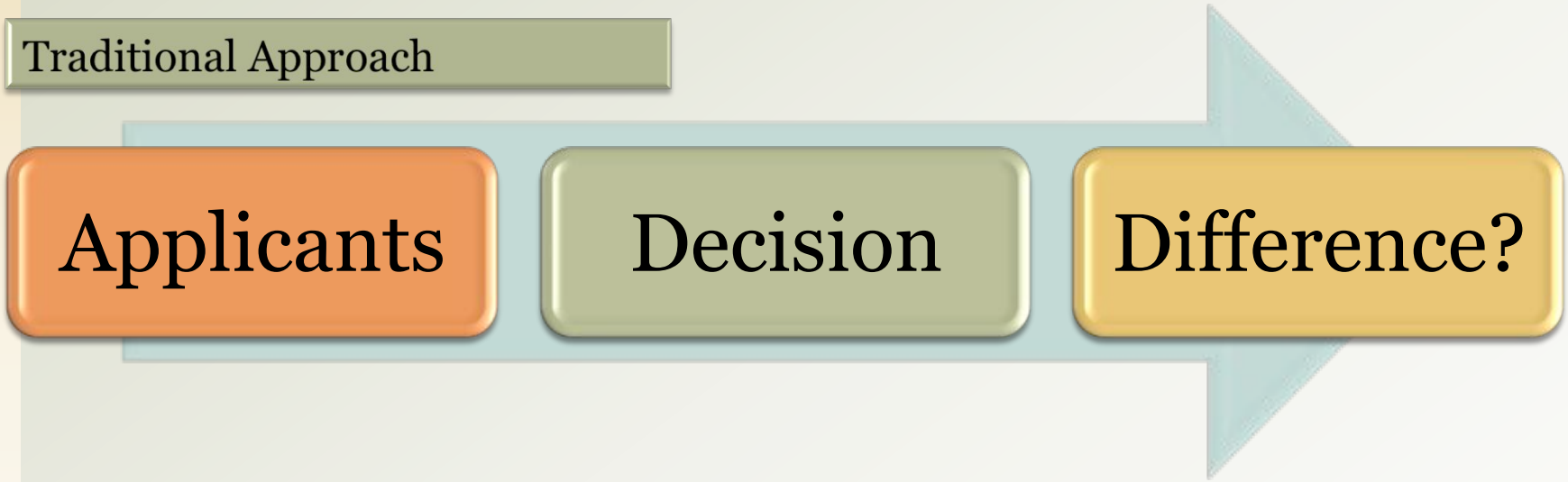
How Binary Logistic Regression Works



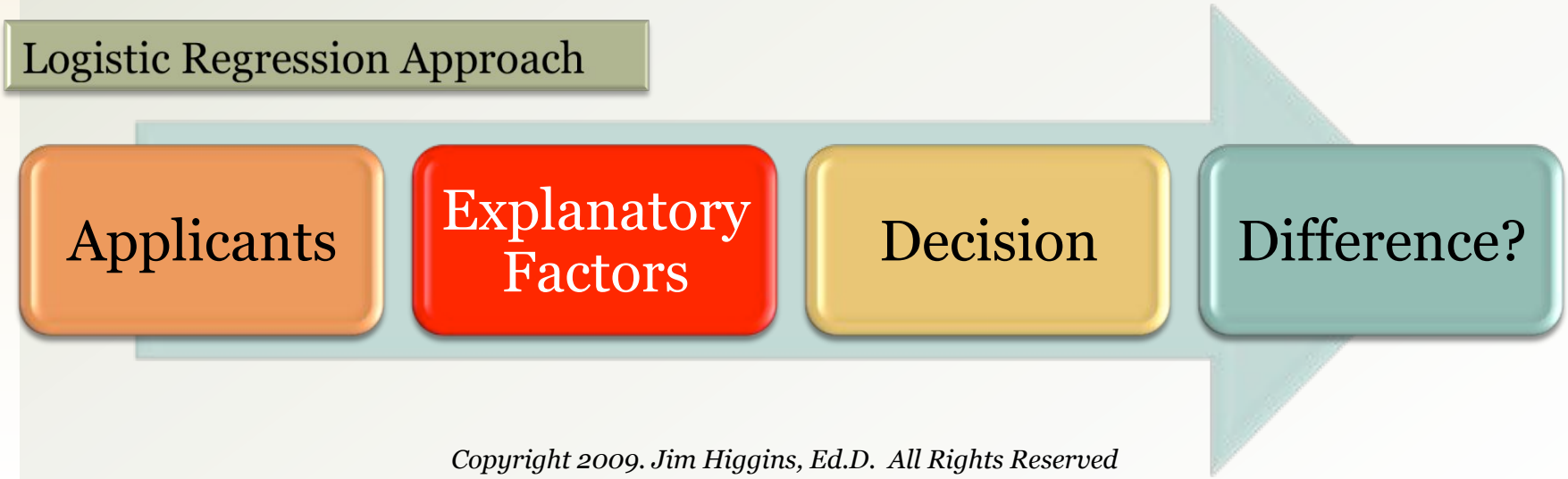
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

The Real Advantage of Binary Logistic Regression

Traditional Approach



Logistic Regression Approach



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

How You Plan To Use Your Test Is Critical!

- **Pass/Fail Cutoffs:**
  - "Normal Expectations of Acceptable Proficiency in the Workplace" (Guidelines, 5H)
  - Modified Angoff (U.S. v. South Carolina, USSC)
- **Banding:**
  - Substantially Equally Qualified Applicants
  - Statistically Driven (use Std. Error of Difference)
- **Ranking:**
  - Is there adequate score dispersion?
  - Does the test have high reliability? (e.g., >.85)
  - Is the KSA performance differentiating?
- **Weighted/combined with other tests**
  - How are the weights related to the job
  - Do they come from the job analysis or SME ratings?

Copyright 2007 Biddle Consulting Group
Used by Permission

How Tests Can Be Used

Applicant	Score
Tom	100
Stacy	100
Bob	100
Frank	100
Julie	99
Rozanne	99
Mark	98
Luke	98
Henry	97
Paul	97
Peter	96
Rebecca	96
Alyssa	95
Matthew	94
John	93
Annette	93
Ray	92
Thomas	91
Julissa	90

- **Ranking** assumes one applicant is *reliably* more qualified than the other
- **Banding** considers the *unreliability* of the test battery and "ties" applicants
- **Pass/fail cutoffs** treat all applicants as either "qualified" or "not qualified"
- **Weighting/combining** test scores can be done using "compensatory" or using cutoff on each test then weighting results

Copyright 2007 Biddle Consulting Group
Used by Permission

Characteristics of Pass/Fail Cut Scores

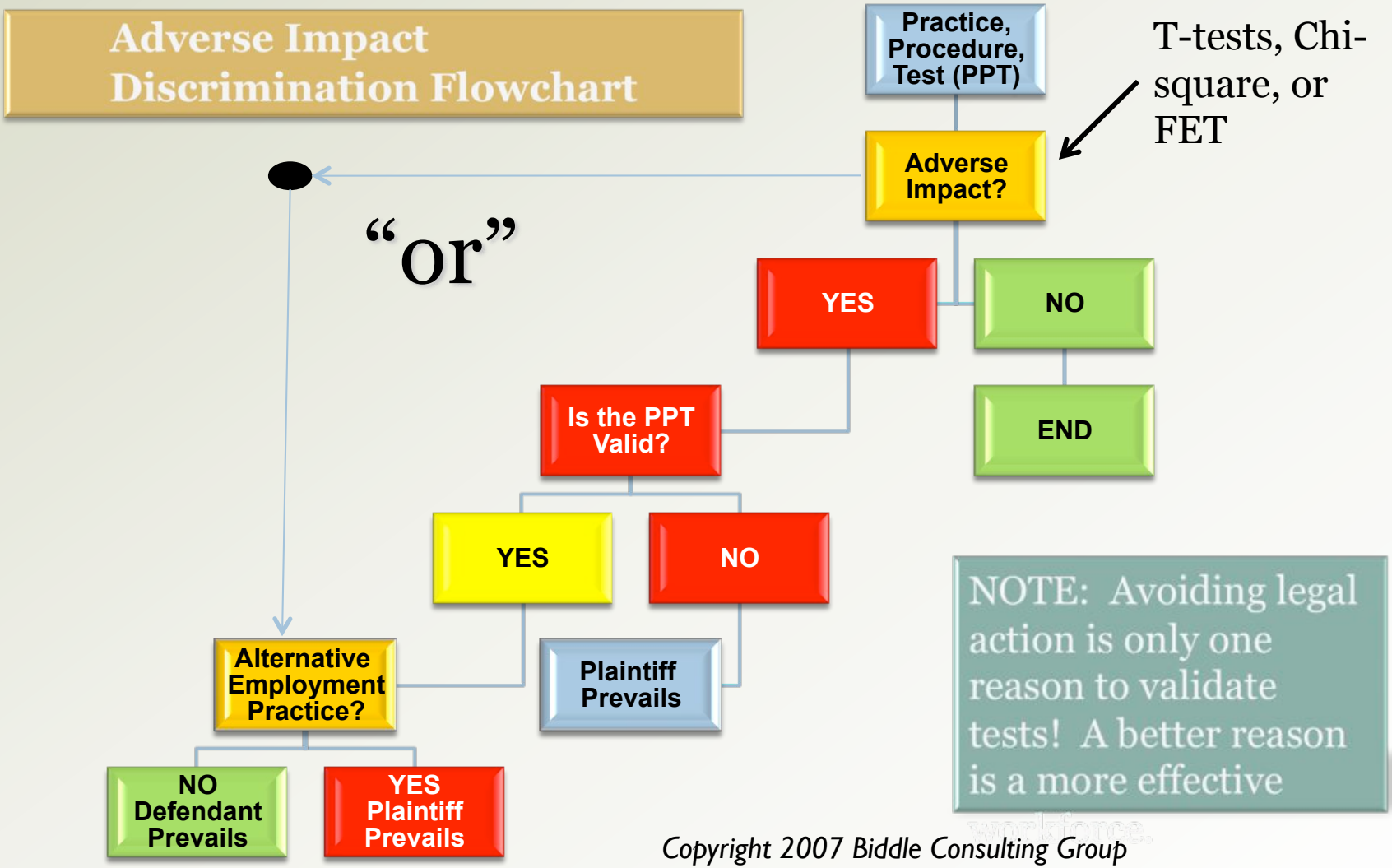
- NOT TYPICALLY DEFENSIBLE WHEN:
  - Using an arbitrary cutoff (e.g., 70%)
  - Using applicant scores to benchmark (e.g., setting cutoff scores at mean-SD of applicant scores)
- TYPICALLY DEFENSIBLE WHEN:
  - Consider "Normal expectations of acceptable proficiency in the workplace" (Guidelines, 5H)
  - Usually requires SME-level data <u>or</u> ratings
  - Tied to job performance
- FACTORS TO CONSIDER:
  - Is the test supported by **content validity information** or **criterion-related information**?
  - How critical are the KSAs measured?
  - Does the test measure "baseline" or "differentiating" KSAs?
  - How would current incumbents perform on this test?

Copyright 2007 Biddle Consulting Group
Used by Permission

Comparison of Score Uses

Factor	Ranking	Banding	Pass/Fail Cutoffs
Validation Requirements	High	Moderate	Low
Adverse Impact	High	Moderate	Low
Defensibility	Low	High	High
Litigation "Red Flag"	High	Moderate	Low
Cost to do so Defensibly	High	Moderate	Low
Applicant Flow	Restrictive/ Controllable	Moderate/ Controllable	High
Development Time to do so Defensibly	High	Moderate	Low

Determining Adverse Impact in Testing



NOTE: Avoiding legal action is only one reason to validate tests! A better reason is a more effective workforce.

Copyright 2007 Biddle Consulting Group
Used by Permission

New You Can Use! Microsoft Excel's Data Analysis ToolPak!

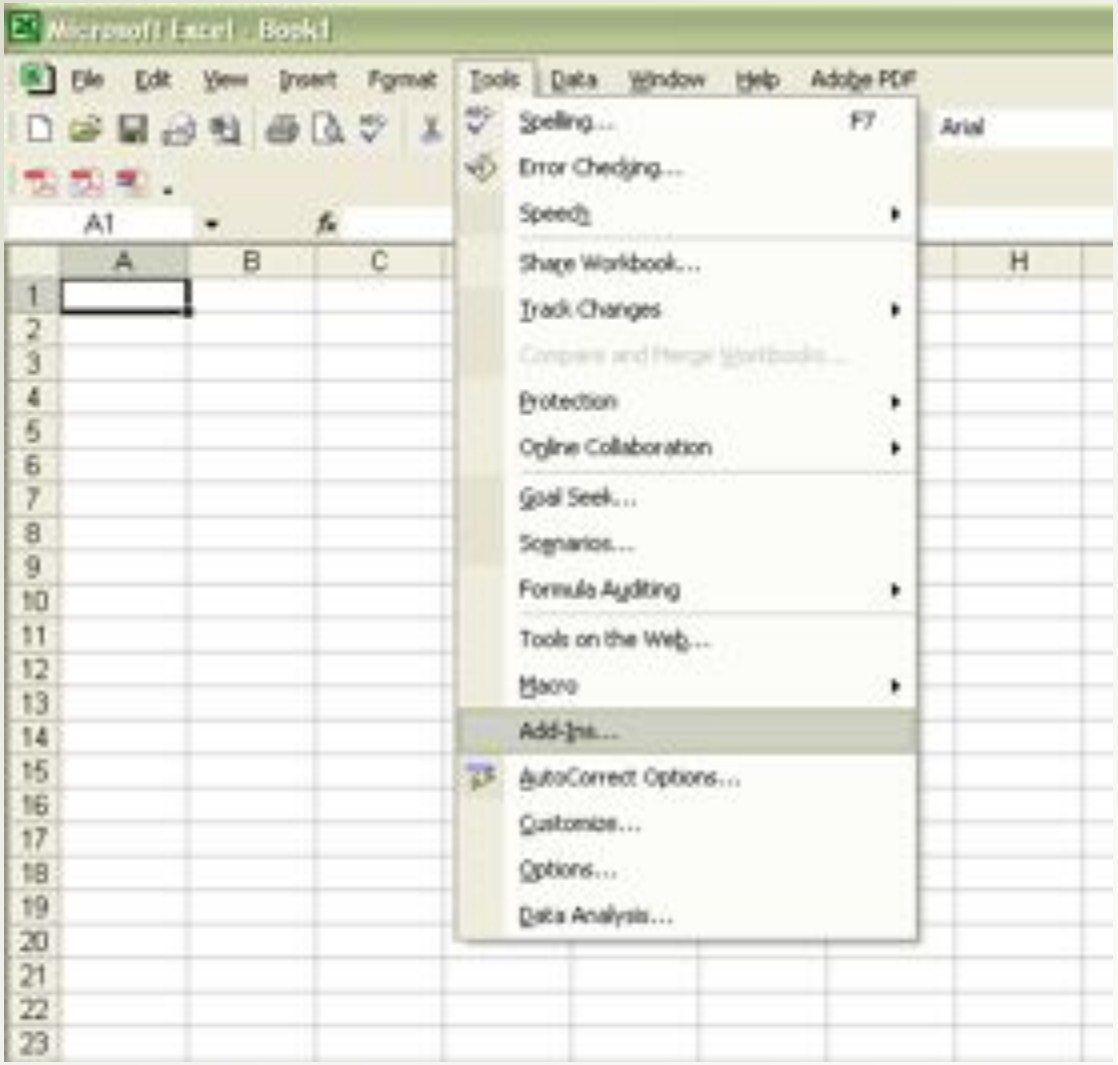
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

How to Run Multiple Regression in Microsoft Excel

- **Step 1** – Install the data analysis toolpak
- **Step 2** – Gather the appropriate data that you believe your organization uses to make compensation decisions
- **Step 3** – Run Compensation Analyses
- **Step 4** – Consider anecdotal data/cohort analysis
- **Step 5** – Calculate amount needed to eliminate statistical significance—if necessary

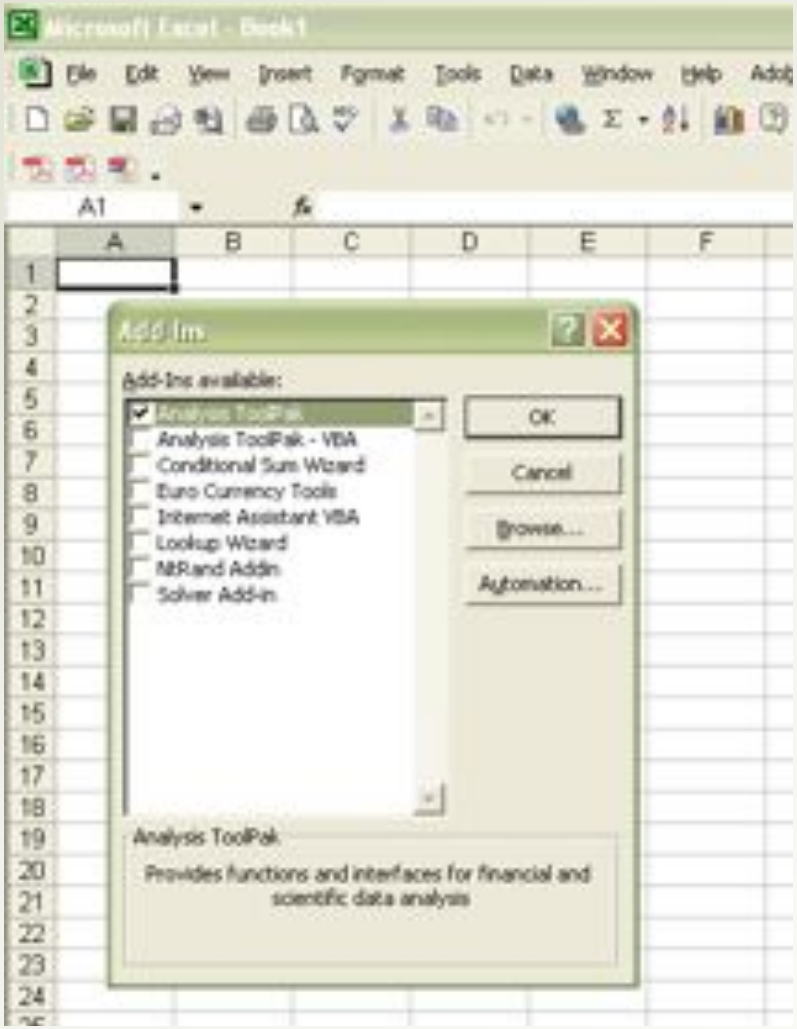
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

How to Install the ToolPak



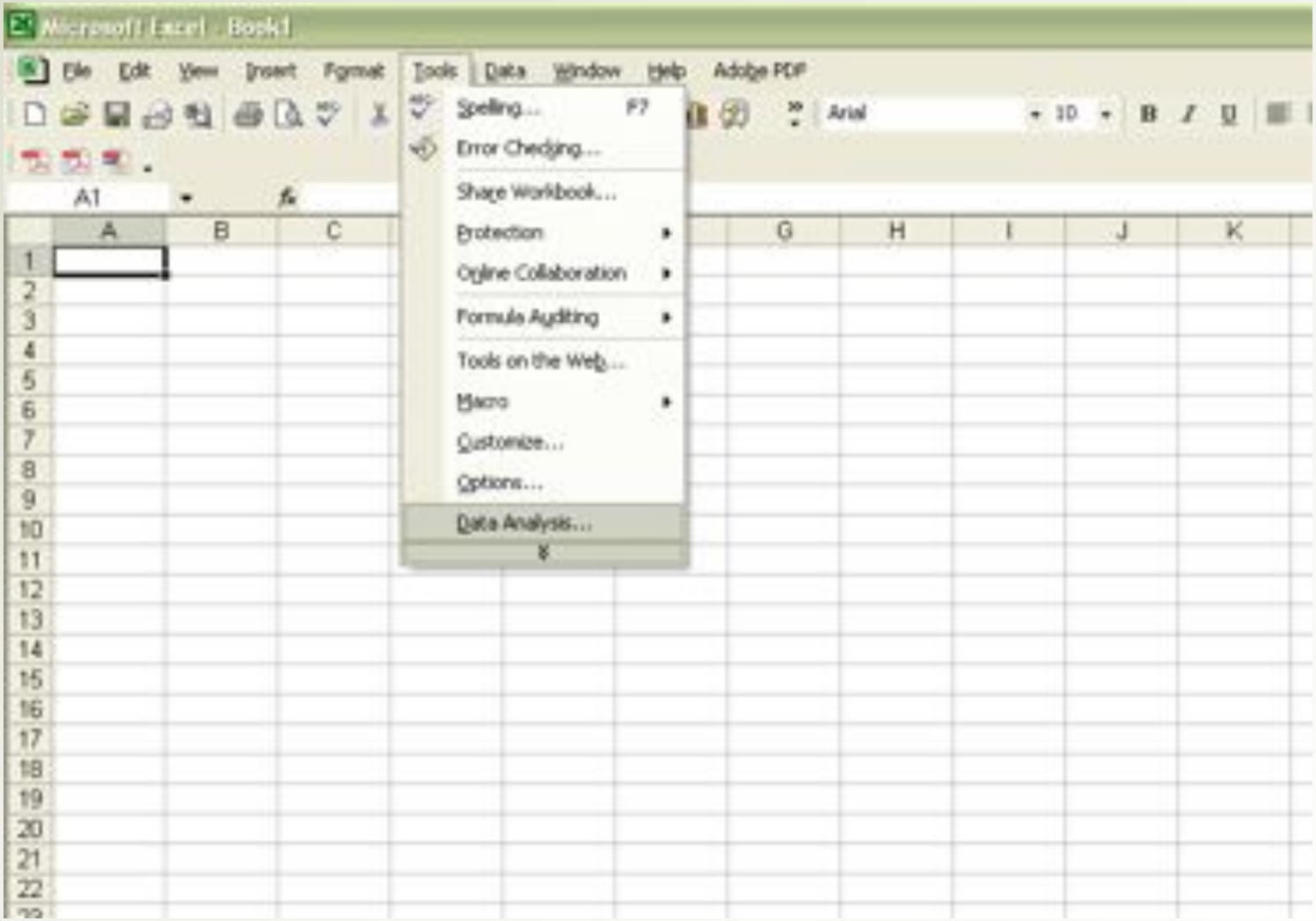
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

How to Install the TookPak



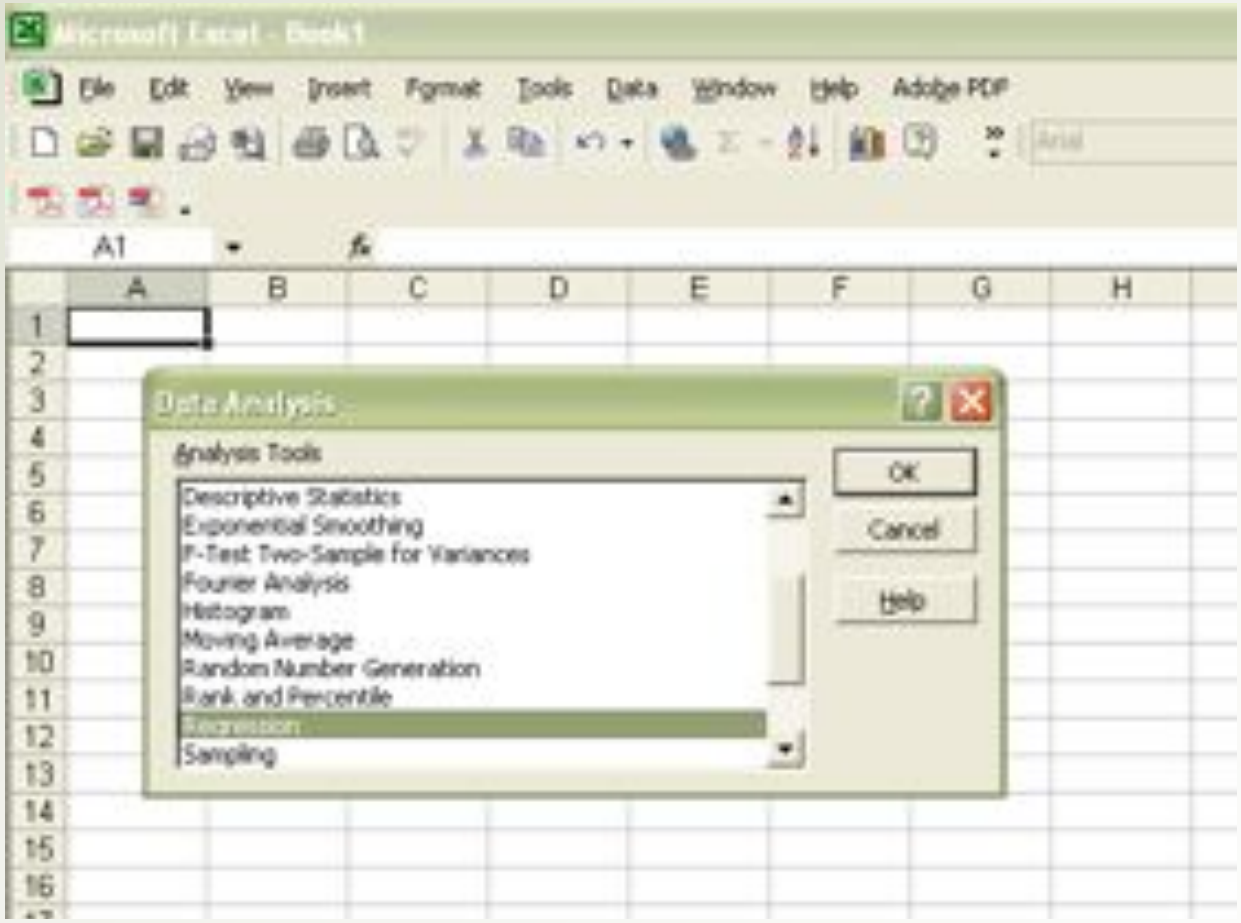
Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

How to Install the TookPak



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

What the ToolPak Does For You



Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

Some Working Examples

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved

For more information

- If you need statistical help…
- Would like to learn more about statistics…
- Have the opportunity to receive free training…
- Become a member of the BCG Institute for Workforce Development online learning community. It is free and is not used to sell any product or service.
- Visit: www.BCGinstitute.org

Copyright 2009. Jim Higgins, Ed.D. All Rights Reserved